ORIGINAL RESEARCH

Pathogenic and Uncertain Genetic Variants Have Clinical Cardiac Correlates in Diverse Biobank Participants

Tess D. Pottinger, MS; Megan J. Puckelwartz, PhD; Lorenzo L. Pesce, PhD; Avery Robinson, BS; Samuel Kearns, BS; Jennifer A. Pacheco, BS; Laura J. Rasmussen-Torvik, PhD; Maureen E. Smith, MS; Rex Chisholm, PhD; Elizabeth M. McNally, MD, PhD

***Background*—**Genome sequencing coupled with electronic heath record data can uncover medically important genetic variation. Interpretation of rare genetic variation and its role in mediating cardiovascular phenotypes is confounded by variants of uncertain significance.

***Methods and Results*—**We analyzed the whole genome sequence of 900 racially and ethnically diverse biobank participants selected from a single US center. Participants were equally divided among European, African, Hispanic, and mixed races/ethnicities. We evaluated the American College of Medical Genetics and Genomics medically actionable list of 59 genes, focusing on the cardiac genes. Variation was interpreted using the most recent reports in ClinVar, a database of medically relevant human variation. We identified 19 individuals with pathogenic or likely pathogenic variants in cardiac actionable genes (2%) and found evidence of related clinical correlates in the electronic health record. Participants of African ancestry, compared with those of European ancestry, had more variants of uncertain significance in the medically actionable genes including the 30 cardiac actionable genes, even when normalized to total variant count per person. Longitudinal measures of left ventricle size from ≈400 biobank participants (1723 patient-years) were correlated with genetic findings. The presence of ≥1 uncertain variant in the actionable cardiac genes and a cardiomyopathy diagnosis correlated with increased left ventricular internal diameter in diastole and in systole. In particular, *MYBPC3* was identified as a gene with excess variants of uncertain significance.

***Conclusions*—**These data indicate that a subset of uncertain genetic variants may confer risk and should not be considered benign. **(*J Am Heart Assoc*. 2020;9:e013808. DOI: 10.1161/JAHA.119.013808.)**

Key Words: biobank • cardiomyopathy • left ventricle • medically actionable genes • variants of uncertain significance

Genetic information is increasingly being used in medical decision making, especially for familial cancers and cardiovascular diseases for which the identification of rare genetic variants can inform care for patients and family members at risk.[1–3] Genetic variants segregating for disease are interpreted as pathogenic or likely pathogenic, and this type of genetic information is diagnostic and useful for clinical management.[4] Variants of uncertain significance (VUSs) are those genetic variants about which information is insufficient to adjudicate a pathogenic or benign classification. The VUS designation often arises for rare or unreported missense variants, and this designation is of low medical utility because its pathogenic status is unknown.[4] To improve the reliability of genetic interpretation, the ClinVar database was developed as an online catalog of genetic variation relevant to human health (https://www.ncbi.nlm.nih.gov/clinvar).[5] Genetic testing laboratories regularly contribute to and update ClinVar's compendium of human health variation. ClinVar, combined with data from large, deidentified, population sequence databases, is enhancing clinical genetic testing interpretation.

The American College of Medical Genetics and Genomics (ACMG) designated 59 genes as having variation that is medically actionable when the variation is classified as *pathogenic* or *likely pathogenic*.[6,7] Not all variation in the ACMG genes is actionable because some variants are found at high population frequency, making their designation benign or likely benign. Uncertain variants are neither pathogenic or benign. However, it can be expected that some of these

From the Center for Genetic Medicine (T.D.P., M.J.P., A.R., S.K., J.A.P., M.E.S., R.C., E.M.M.), Departments of Pharmacology (M.J.P.), Preventive Medicine (L.J.R.-T.), and Cell and Molecular Biology (R.C.), Northwestern University Feinberg School of Medicine, Chicago, IL; Computation Institute, University of Chicago, Chicago, IL (L.L.P.).

Accompanying Tables S1 through S9 and Figures S1 through S5 are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.013808

Correspondence to: Megan J. Puckelwartz, PhD, and Elizabeth M. McNally, MD, PhD, Center for Genetic Medicine, 303 E. Superior St., SQ 5-511, Chicago IL 60611. E-mails: m.puckelwartz@northwestern.edu; elizabeth.mcnally@northwestern.edu

Received July 2, 2019; accepted December 4, 2019.

Clinical Perspective

What Is New?

- Genetic variants of uncertain significance are increased in the cardiac actionable genes in biobank participants of African ancestry compared with those of European ancestry.
- Variants of uncertain significance in cardiac actionable genes associate with changes in left ventricular dimensions over time; therefore, variants of uncertain significance should not be considered benign because they may confer risk for cardiovascular disorders.

What Are the Clinical Implications?

- Returning information on variants of uncertain significant to biobank participants should be considered, especially when clinical correlates are present, because these genetic variants may help direct clinical risk reduction for patients and their family members.

uncertain variants are, in fact, medically important. It has been recommended that known pathogenic and likely pathogenic results be returned for the actionable genes, even for biobank participants.[8–10] VUSs are not typically reported and returned to biobank participants because the risks associated with these variants have not been determined. Race and population diversity influence the interpretability of genetic testing results.[11–13]

We used whole genome sequencing (WGS) on a diverse cohort of biobank participants from a single metropolitan site in the United States. We assessed genetic variation across self-reported race/ethnicity groups, focusing on medically actionable genes and variants previously reported in ClinVar. We found that participants of African ancestry had a significantly greater number of VUSs compared with participants of European ancestry. This increase in uncertain variants was present across all genes including the ACMG medically actionable genes and the actionable cardiac genes. We assessed the prevalence of VUSs across this diverse population and examined the association of VUSs with echocardiographic indicators of cardiac pathology. Uncertain variants in the cardiac medically actionable genes were significantly associated with changes in left ventricular (LV) dimensions. These data underscore the complexities of variant interpretation, especially in a diverse population.

Methods

Transparency and Openness Statement

Genomic data are available through dbGaP accession number phs001191.v1.p1. Because of the sensitive nature of the clinical data collected for this study, requests to access the data set by qualified researchers trained in human subject confidentiality protocols may be sent to Maureen Smith or Elizabeth McNally at Northwestern University.

Study Approvals

All participants provided written consent for participation in NUgene, and this work was performed under the ethics and regulatory approval of Northwestern University's institutional review board (STU0010003).

NUgene Cohort

The NUgene biobank includes adult participants who receive care at Northwestern Medicine. Inclusion and exclusion criteria for participation in the NUgene biobank have been described.[14] Through its participation in the eMERGE (Electronic Medical Records and Genomics) consortium,[15,16] the NUgene biobank selected 900 participants for WGS. Race/ethnicity was collected using 2 methods. Approximately 25% of participants used a questionnaire in which race/ethnicity was selected based on multiple general categories including African, Asian (East and South), European, Hispanic, Pacific Islander, Middle Eastern, and Native American ancestry. Participants who completed this questionnaire could select ≥1 category. The remaining 75% of patients were classified based on grandparent ancestry using the same general categories noted. Because the numbers of individuals who identified as Asian, Asian Indian, Middle Eastern, and Pacific Islander through both methods were low (<1% of the sample), they were reclassified as *Other* and excluded from further analysis. Individuals who self-reported as multiracial were classified as *mixed*, and individuals with at least 1 grandparent who did not classify similarly to the other 3 grandparents were classified as mixed. For instance, if an individual selected 3 grandparents as European ancestry and 1 grandparent of African ancestry, this individual was classified as mixed.

Whole Genome Sequencing

WGS was performed on an Illumina XTen machine at the Genome Center at Washington University School of Medicine, yielding >100 GB of data per sample. This depth correlates with >30-fold coverage across the genome, providing more even coverage across both noncoding and coding intervals. WGS data from the NUgene cohort were aligned to the human genome reference sequence GRCh37/hg19 using the Burrows-Wheeler Aligner (BWA). Variants were called using the Genome Analysis Tool Kit (GATK v3.3.0).[17,18] These analyses were conducted using the MegaSeq Pipeline.[19]

Table 1. Demographic Characteristics of NUgene Cohort by Self-Reported Race/Ethnicity

	NUgene	African	European	Hispanic	Mixed
Participants, n	886	235	206	233	212
Age, y, mean (SD)	52 (12)	55 (12)	50 (11)	52 (11)	51 (12)
Sex, male, %	34	31	55	31	20

Population Structure

The first 2 components generated by principal component analysis (PCA) were used to estimate global ancestry in the NUgene 900 cohort. PCA was conducted using singular-value decomposition of shared variants in the NUgene cohort with ≈5 million biallelic variants distributed across the genome. Coding and noncoding variants were identified using ANNOVAR.[20] Self-reported race was compared with the ancestry groupings determined by the PCA of all variants. An identity-by-state analysis was conducted to identify first-degree relatives within each homogeneous cluster to reduce overfitting in subsequent analytical models. To carry out the identity-by-state analysis, 5 individuals were removed from the European cluster, 1 from the African cluster, and 4 from the Hispanic cluster before this analysis to ensure homogeneous race/ethnicity clusters based on visual assessment of PCA graphs to classify large outliers. Analyses were carried out using PLINK v1.9 and R v3.5.1.

ClinVar Analysis

Nonsynonymous coding variants were queried and indexed per genome using ClinVar (February 11, 2019, adjudication of variants from the variant call format (VCF) file).[5] These analyses included designations of *likely pathogenic, pathogenic, pathogenic/likely pathogenic, uncertain significance*, and *not reported* in ClinVar. Variants were not further filtered after ClinVar designation; therefore, variants that are not reported in ClinVar are likely ultra-rare. The ClinVar VCF file used to annotate these variants included only the most recent designation, as determined by February 11, 2019. Generalized regression models adjusting for age and sex were used to determine overall differences and pairwise differences between self-reported racial/ethnic groups. To correct for differences in the proportion of variants among the self-reported race/ethnicity groups, variants were normalized based on total variant counts per person (an individual's variant count was divided by the individual's total number of variants). Analysis type was determined by the distribution of the outcome variables. To correct for multiple testing for the pairwise analysis comparisons, we used the Bonferroni

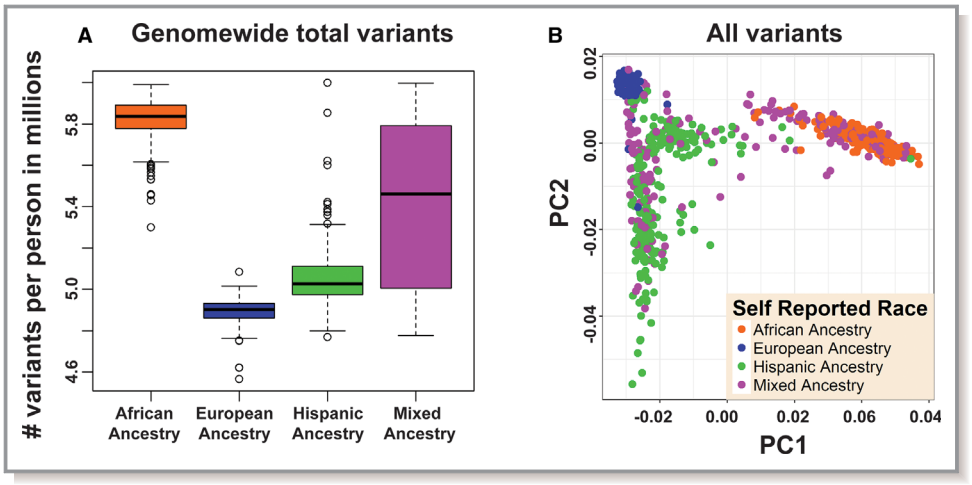


Figure 1. Variant number per person in the NUgene biobank genomes. Biobank participants of African ancestry had significantly greater genetic variation than those of European ancestry. **A**, The average number of total variants across the genome is shown for each group based on self-reported race; African ancestry (5 815 632±105 545 variants per individual genome), European ancestry (4 891 014±64 442 variants per individual genome, Hispanic ancestry (5 056 797±151 424 variants per individual genome), and mixed (5 407 116±392 139 variants per individual genome; $P<0.0001$, ANOVA across all groups). **B**, Shown are results from principal component analysis using singular-value decomposition of shared genetic variants among biobank participants; the first two principal components are shown (PC1 and PC2). Participants of self-reported African and Hispanic ancestry displayed a more heterogeneous genetic pattern than those of European and mixed ancestry for all variants. Self-reported race/ethnicity (colored circles) are displayed on a genetic clustering background and derived from self-report and grandparent race/ethnicity.

method for correction using the R software package.[21] For this analysis, 6 tests were used for correction.

Echocardiogram Analysis With VUSs

Echocardiogram, ECG, and demographic data were queried from the electronic health record, and individual measures were obtained for LV internal diameter–diastole (LVIDd), LV internal diameter–systole, interventricular septal end-diastole, and LV ejection fraction gathered before 2017 and spanning as much as 14 years of data. We tested for relatedness of individuals with echocardiography measures. There were 2 individuals with a diagnosis of cardiomyopathy who had second-degree relatives; however, none of these relatives have VUSs in the cardiac medically actionable genes and thus are not included in the following analysis. The association of longitudinal echocardiogram data and the count of VUSs in the cardiac medically actionable genes was calculated. These counts were coded as 0 and ≥1 to reflect the number of variants found in the cardiac actionable genes (because variant counts of 2 were found in only 22 individuals, these counts were recorded as ≥1). Some participants had multiple echo measurements in a given year. For these participants, the median value was used for analysis. A longitudinal model that controlled for age at echo measurement, sex, and self-reported race/ethnicity with an unstructured covariance matrix was used for analysis with the assumption that missing values were missing at random. Year was used as the time component for this analysis. Race/ethnicity-specific analyses were also conducted using a similar model. A sensitivity analysis was conducted, as above, for individuals who had an *International Classification of Diseases, Ninth Revision* (*ICD-9*) code for cardiomyopathy (*ICD-9* code 425, all subcodes). The longitudinal echocardiogram data and the count of variants not identified in the February 2019 ClinVar database (referred to in this study as *unreported*) in the cardiac medically actionable genes were also analyzed using a similar method. Analyses were completed using SAS 9.4 (SAS Institute) and R v3.5.1 (R Foundation for Statistical Computing).

Results

Genetic Diversity in a US Metropolitan Healthcare Cohort

Through its participation in the eMERGE consortium,[15,16] the NUgene biobank selected 900 participants for WGS. The NUgene biobank represents a medical biobank in that the participants receive health care at a single institution, and their healthcare data are stored in the Enterprise Data Warehouse at Northwestern Medicine. Overall, 23% of the sequenced cohort (n=210) was selected based on having diagnostic and/or procedure codes or medications indicating 1 of 4 conditions of interest to eMERGE network investigators: 95 with atopic dermatitis (*ICD-9* codes 691.8-692.9 and medication codes), 118 with cancer (*ICD-9* codes 173–209), 56 with cardiomyopathy (*ICD-9* codes 425.1 and 425.4; and echocardiogram with both ejection fraction and other LV measures present), and 180 with chronic rhinosinusitis.[22] Of the 210 patients, 180 had >1 of these 4 diagnoses and 30 had only 1 of the 4 diagnoses. Seventy-seven percent of the cohort (n=690) was selected only for race/ethnicity and without regard to clinical diagnostic codes. WGS was applied to these 900 diverse individuals; sequencing reads were aligned to the human genome reference sequence GRCh37/hg19, and variants were called using the MegaSeq Pipeline, which utilized BWA and GATK best practices. Of the 900 genomes, 5 were excluded because of sampling error identified through sex mismatch and/or possible sample contamination. The remaining 895 genomes distributed as follows based on self-reported race/ethnicity: African, 26%; European, 23%; Hispanic, 26%; mixed, 24%; and other, 1%. Because the sample size for individuals who identified as other was small, they were excluded from this analysis

Table 2. Genetic Variation in the NUgene Cohort Classified by ClinVar

	No. of People	Mean/ Person	Range
Pathogenic variants			
All genes	886	6.51	1 to 19
Medically actionable genes	23	0.026	0 to 1
Cardiac actionable genes	10	0.011	0 to 1
Pathogenic/likely pathogenic variants			
All genes	173	0.22	0 to 4
Medically actionable genes	13	0.014	0 to 1
Cardiac actionable genes	5	0.0056	0 to 1
Likely pathogenic variants			
All genes	660	1.02	0 to 4
Medically actionable genes	7	0.0079	0 to 1
Cardiac actionable genes	4	0.0045	0 to 1
VUS			
All genes	886	33.42	15 to 58
Medically actionable genes	385	0.58	0 to 4
Cardiac actionable genes	191	0.24	0 to 3
Unreported variants			
All genes	886	13 430	10 042 to 15 282
Medically actionable genes	886	2.57	2 to 13
Cardiac actionable genes	886	2.24	2 to 6

VUS indicates variants of uncertain significance.

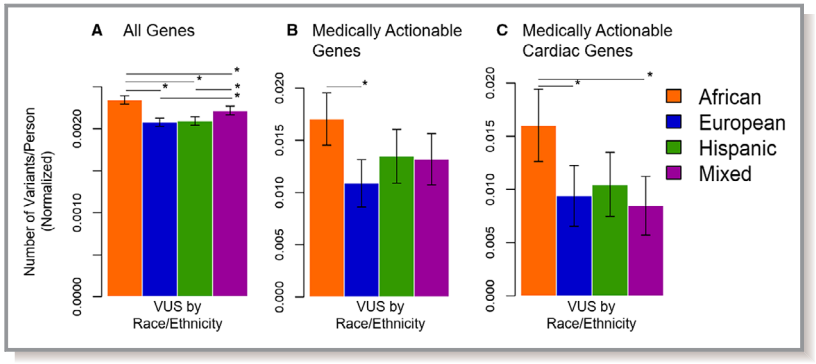


Figure 2. Biobank participants of African ancestry have significantly more variants of uncertain significance (VUSs) than other groups. **A**, When evaluating all coding genes, VUS count per person, as determined by most recent ClinVar report, was greater in biobank participants of African ancestry compared with other groups (P<0.0001, ANOVA across all groups). **B**, When evaluating the 59 medically actionable genes, biobank participants of African ancestry had more VUSs per person than those of European ancestry (P<0.0001) but did not differ from other groups. **C**, When evaluating the 30 cardiac actionable genes, biobank participants of African ancestry had more VUSs per person than the European and mixed ancestry groups (P<0.0001) but not the Hispanic ancestry group. Pairwise-comparison exact P values are shown in Table S2 and are adjusted for multiple comparisons.

(Table 1). Genetic variation correlated with self-reported race/ethnicity. Individuals of European ancestry had the lowest number of variants per person compared with those in the other groups (4.9 million per person for European compared with 5.8 million per person for African, P<0.05; Figure 1A). Racially mixed and Hispanic individuals had variant counts of 5 million and 5.4 million per person, respectively—between those of African and European ancestry. An identity-by-state analysis was performed, and 2 pairs of first-degree relatives and 4 pairs of second-degree relatives were identified with this analysis.

Genomic data were annotated for noncoding and coding variation. As expected, >99% of all genetic variants were noncoding. Considering the nonsynonymous coding variants, the majority were missense variants (79%), and a smaller percentage were stop/loss gain (1.0%), small in-frame insertion/deletions (2.5%), frameshifts (2.5%) and splice sites (15%; Figure S1). The number of variants observed only once in the entire data set, a measure of rare variation, was highest in participants of African ancestry (52 865±8320). Participants of European ancestry had the least of any group (30 217±3992, P<0.05, ANOVA across all groups; Figure S2).

When considering genetic variation across the genome, those who identified as being of European ancestry were tightly clustered on the PCA plot (Figure 1B, blue). In contrast, those who identified as having African ancestry clustered less tightly (Figure 1B, orange). Notably, this group had African and European admixture with as much as 50% European ancestry. Those identifying as Hispanic had a nonuniform structure with 2 major groupings, including one that clustered along the European–Asian line, and the other without a clear racial grouping, indicating more mixed ancestry (Figure 1B, green).

Participants of African Ancestry Have More VUSs

Genetic testing is increasingly used in adult healthcare settings, but the interpretation of genetic results is complicated by the rare frequency of many genetic variants.[23,24] Genetic testing is further complicated by the observations that pathogenic and likely pathogenic variants are found at higher frequencies than the diseases specified by these variants.[25,26] We queried the number of nonsynonymous variants in the NUgene biobank genomes that were previously reported in ClinVar, a database of clinically relevant genetic information[5] (Table 2). The most recent variant adjudication was used for this analysis; ≈90% of variants were adjudicated since 2015, the time when the ACMG guidelines were released.[6] People of African ancestry are known to have greater genetic variation than either their European or Hispanic counterparts. Therefore, variant counts were divided by total variant count per person to account for this baseline difference across populations.[27] NUgene participants of African ancestry had more VUSs than participants of European, Hispanic, and mixed ancestry, even after normalization to the total number of variants per person (P<0.0001; Figure 2A).

The NUgene genomes were then queried for nonsynonymous variation in the 59 medically actionable genes (Figure 2B).[7] The mean normalized number of pathogenic

Table 3. Pathogenic/Likely Pathogenic Variants in Cardiac Actionable Genes in the NUgene Cohort and Corresponding Electronic Health Record Data

Type	Gene	Disease Linked to Gene	Sex	Age, y	Race/Ethnicity	Variant	Relevant Phenotype
P	*KCNQ1*	Long QT syndrome	F	71	African	p.Val205Met	Max QTc: 534 ms
P	*KCNQ1*	Long QT syndrome	F	37	Hispanic	p.Ile198Val	Max QTc: 485 ms
P/LP	*KCNH2*	Long QT syndrome	F	61	Hispanic	p.Arg948His	Max QTc: 479 ms
LP	*KCNQ1*	Long QT syndrome	F	48	African	p.Glu146Gly	Max QTc: 475 ms
P	*PKP2*	ARVC	F	30	Mixed	p.Tyr130Ter	Max QTc: 476 ms
P	*DSC2*	ARVC	F	65	European	p.Tyr332Ter	*
P/LP	*LDLR*	Hypercholesterolemia	F	59	Hispanic	p.Gly592Glu	Max total CHOL: 308 mg/dL
LP	*LDLR*	Hypercholesterolemia	F	57	African	p.Ser648Ala	Max total CHOL: 251 mg/dL
P/LP	*LDLR*	Hypercholesterolemia	F	31	Mixed	p.Cys681Ter	Max total CHOL: 249 mg/dL
LP	*PCSK9*	Hypercholesterolemia	M	67	African	p.Asp204Asn	Max total CHOL: 243 mg/dL
P	*PCSK9*	Hypocholesterolemia†	M	49	African	p.Tyr142Ter	. . .
P	*APOB*	Hypercholesterolemia	F	49	Mixed	p.Tyr4343Cysfs	. . .
P	*MYBPC3*	HCM	M	38	European	p.Trp792Valfs	Max LVPWD: 1.8 cm
P/LP	*MYBPC3*	HCM	F	55	Mixed	p.Glu1096Ter	Max LVPWD: 1.3 cm
P	*MYBPC3*	HCM	F	51	European	p.Trp792Valfs	Max LVPWD: 1.1 cm
P	*MYBPC3*	HCM	F	37	European	c.1928-2A>G	. . .
P	*MYL2*	HCM	F	32	European	p.Pro95Ala	. . .
P/LP	*SCN5A*	Brugada syndrome	F	59	Mixed	p.Val845Cysfs	Max QTc: 526 ms
LP	*SCN5A*	Brugada syndrome	M	43	European	p.Ser1135Ile	. . .

Ellipses indicate test not conducted. ARVC indicates arrhythmogenic cardiomyopathy; CHOL, cholesterol; F, female; HCM, hypertrophic cardiomyopathy; LP, likely pathogenic; M, male; Max, maximum; P, pathogenic.
*Values not available.
†Variant linked to hypocholesterolemia.

and likely pathogenic variants in the 59 medically actionable gene lists did not differ across groups; however, the total number of these variants was very small. Among the medically actionable genes, individuals of African ancestry were more likely to have VUSs than individuals of European ancestry, even after normalization to total variant count per person ($P<0.05$; Figure 2B). Numbers of VUSs were similar between the other ancestry groups (Figure 2B). The medically actionable genes were subdivided into cardiac and cancer genes and similarly analyzed (Table S1). For cardiac medically actionable genes, participants of African ancestry had more VUSs than European and mixed individuals, even after normalization to total variant count ($P<0.05$; Figure 2C, Table S2).

VUSs for Medically Actionable Cardiac Genes Correlate With LV Measures

Of the 59 medically actionable genes, 30 are linked to cardiovascular conditions (Table S1). Nineteen participants had ClinVar-adjudicated pathogenic and likely pathogenic variants in these genes, with no single participant having >1 pathogenic and likely pathogenic variant (Table 3). Of these 19 individuals, 13 had an ECG, echocardiogram, or test of their cholesterol level in the electronic health record, and those with electronic health cardiac data were 10 years older than those without electronic health cardiac information (aged 52 versus 42 years). All 4 individuals who had pathogenic and likely pathogenic variants in long-QT syndrome genes had a prolonged corrected QT interval (QTc) (Table 3), but none carried an *ICD* diagnostic code for long-QT syndrome. Of these 4 individuals, 1 had an *ICD* code for cardiomegaly and myocardial infarction, 1 had an *ICD* code for premature atrial contraction, and 1 had an *ICD* code for atrial fibrillation. The fourth individual did not have any cardiac *ICD* code; however, this individual was <50 years old.

Because a proportion of VUSs are likely to confer phenotype, VUSs within the medically actionable cardiac genes were analyzed for their association with echocardiographic measures. Of the 385 individuals with echocardiographic data in the electronic health records, 108 individuals had at least 1 VUS in the cardiac genes. Of the 30 genes studied, *MYBPC3* had the most VUSs (Figure 3). To determine whether VUS count was associated with cardiac phenotype, a Loess plot was used to

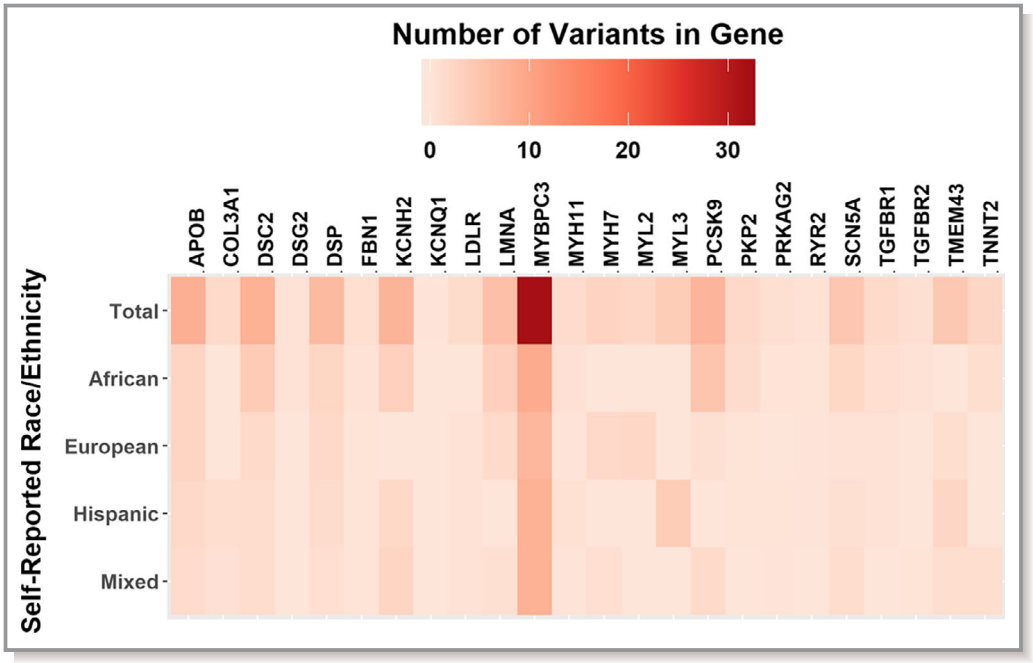


Figure 3. Number of variants of uncertain significance (VUSs) in medically actionable cardiac genes. VUS number in the cardiac actionable genes is indicated across the NUgene cohort (total cohort represented in the top line) and by self-reported race/ethnicity in each line below. VUS count was normalized for gene length.

compare LV measures over time and then correlated with presence of VUSs per person. Longitudinal measures of LVIDd and LV internal diameter–systole increased with VUS count (Figure S3). Longitudinal analyses revealed a significant increase in LVIDd over time in those with ≥1 VUS per person compared with those without VUSs in cardiac genes, and this significance was observed after controlling for age at echocardiogram, sex, and self-reported race/ethnicity ($P<0.05$; Table S3). When evaluating these same data within each race/ethnic group, a similar trend was seen for LVIDd and VUS count for participants of European and Hispanic ancestry. Individuals with cardiac VUSs also showed a similar effect for systolic measurements (LV internal diameter–systole) over time ($P<0.01$; Table S3). In this case, the race/ethnicity-specific analysis identified those of African and European ancestry as having a significant change over time ($P<0.05$). LV ejection fraction and interventricular septal end diameter–diastole showed no significant differences in the pooled analyses. However, in the race/ethnicity analysis, participants of African ancestry with VUSs had a significant decrease in LV ejection fraction over time ($P<0.05$; Table S3). These data indicate that the presence of VUSs within cardiac actionable genes is associated with a change in cardiac dimensions over time.

To determine whether these trends were driven by participants with diagnosed cardiomyopathy, the electronic health record was queried for *ICD-9* codes for cardiomyopathy including *ICD-9* code 425. Ninety-four subjects were identified, including the original 56 preselected at the time of sequencing. For these 94 cardiomyopathy subjects, the association of longitudinal LV dimensions and VUS count per person appeared similar to that of the entire cohort (Figure 4 and Table 4). These trends were similar when either genetic or self-reported race/ethnicity was used for this analysis (Table S4). None of these 94 individuals had a known pathogenic variant in the cardiac medically actionable gene list. Only 24 had VUSs, with only 2 individuals having ≥1 VUS. Because there are more cardiomyopathy genes beyond those on the medically actionable list, we queried variation in 102 cardiomyopathy genes; this list of cardiomyopathy genes was derived from gene panels used in commercial testing laboratories (Table S5). Only 1 individual had a cardiomyopathy pathogenic variant (*TTR* V122I) and a VUS in the cardiac medically actionable genes. Subjects with pathogenic variants lacked VUSs in the cardiac medically actionable genes and thus did not contribute to changes in LV dimensions seen in Figure 4. Twenty-four cardiomyopathy-diagnosed individuals harbored VUSs in cardiomyopathy genes, and these individuals are responsible for the change in LV dimensions over time seen in Figure 4. The list of these VUSs is shown in Tables S6 and S7. These variants were all adjudicated after 2015, making them compliant with current variant adjudication guidelines.[4] Five of 24 participants had VUSs in *MYBPC3*. These data suggest that VUSs contribute to changes in ventricular dimensions and support the importance of interpreting variants in the context of phenotype, where prior probability differs from the general population. All 5 of these

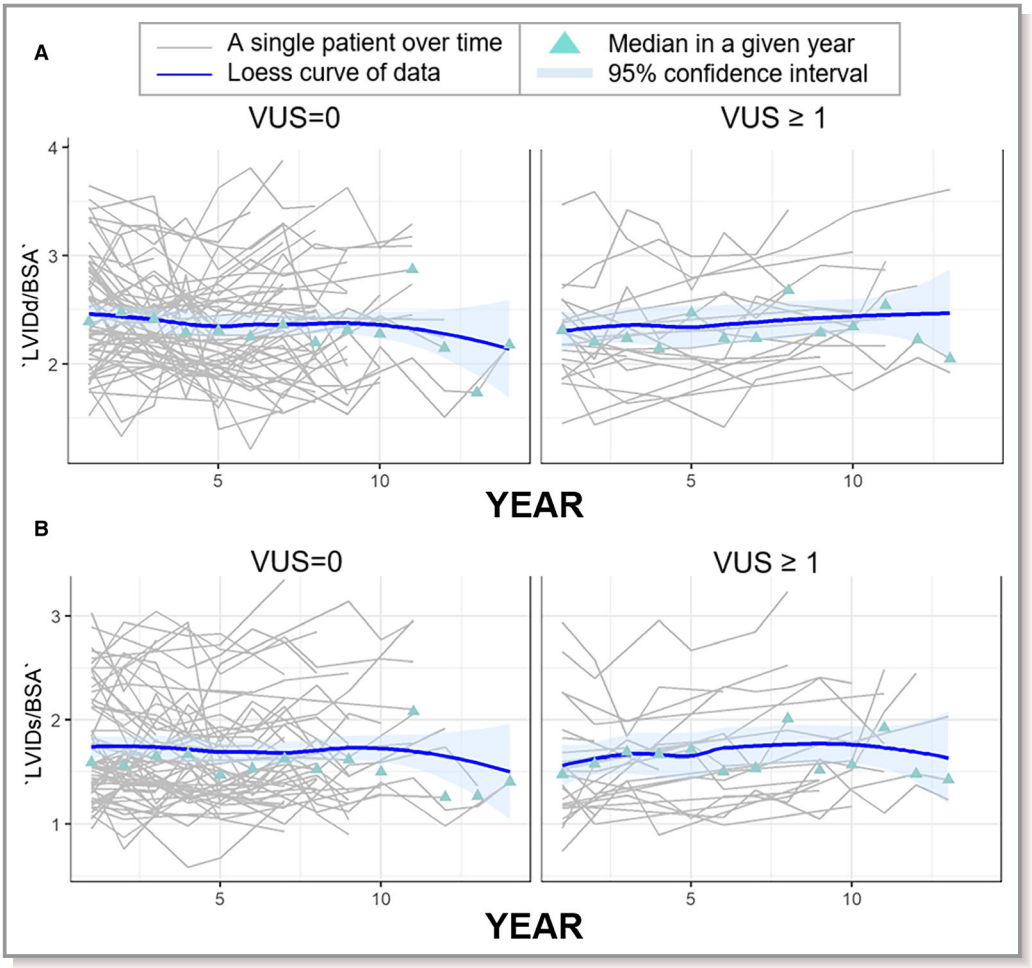


Figure 4. Median left ventricular dimensions correlated with number of variants of uncertain significance (VUSs) in the cardiac actionable genes in 94 patients with a cardiomyopathy diagnosis. **A**, Left ventricular internal diameter in diastole (LVIDd) corrected for body surface area (BSA) over 14 years of echocardiographic data derived from the electronic health record. Those with a VUS or VUSs had an increase in LVIDd, corrected for BSA. **B**, Data for left ventricular internal diameter in systole (LVIDs) corrected for BSA, gathered over the same time frame as (**A**). *P* values for slope difference estimates are shown in Table 4.

variants are predicted to be probably damaging or possibly damaging, using the in silico tool PolyPhen-2.

Evaluating Variants Unreported in ClinVar From Diverse Biobank Participants

We next examined variants not previously reported in ClinVar. Variants were not filtered for frequency, in silico pathogenicity prediction, or potential functional consequence. Across all genes, participants of African ancestry had a greater number of variants not previously reported in ClinVar ($P<0.0001$; Figure S4A). For the medically actionable genes, similar numbers of variants were not previously reported in ClinVar among all groups when variant counts were normalized to the total number of variants per person found in these genes (Figure S4B). For cardiac actionable genes, participants of African ancestry had fewer unreported variants compared with individuals of European and Hispanic ancestry when normalized to the total number of variants found in these genes ($P<0.0001$ for both; Figure S4C and Table S8).

We evaluated the relationship between echocardiographic measurements and unreported variants in the cardiac medically actionable genes (Table 2). The genes with the most unreported variants were *APOB*, *MYH11*, and *DSP* (Figure S5). Longitudinal analyses of LVIDd, LV internal diameter–systole, interventricular septal end-diastole, and LV ejection fraction, corrected for body surface area, showed no difference over time when evaluating unreported variants (Table S9). These analyses controlled for age at echocardiogram, sex, and self-reported race/ethnicity (Table S9). Although there may be variants of clinical impact in this data set, the signal may be masked by the large number of benign variants.

Table 4. Longitudinal Association of LV Measures with VUS Count in Cardiac Actionable Genes from 94 Participants With Cardiomyopathy Diagnostic Codes

Echo Measure VUS ≥1 vs VUS 0	Total Observations	African	European	Hispanic	Mixed
LVIDd/BSA, n	447	99	238	42	68
Slope difference estimate	+0.035	+0.0012	+0.038	+0.015	. . .
P value	0.0035*	0.45	0.044*	<0.0001*	. . .
LVIDs/BSA, n	442	94	239	41	68
Slope difference estimate	+0.047	+0.047	+0.05	. . .	−0.0035
P Value	0.0003*	0.0063*	0.023*	. . .	0.92
LVEF, n	426	90	225	41	70
Slope difference estimate	−0.42	−1.93	−0.25	. . .	+2.19
P Value	0.37	0.007*	0.68	. . .	0.13
IVSDd/BSA, n	465	101	248	45	71
Slope difference estimate	−0.0014	−0.013	+0.008	−0.016	+0.0065
P Value	0.72	0.054	0.17	0.16	0.48

Slope difference estimates from patients having ≥1 VUS to those patients having 0 VUS, and the model controlled for age at echocardiogram, sex, and self-reported race/ethnicity. BSA indicates body surface area; IVSDd, interventricular septal end-diastole; LV, left ventricular; LVEF, left ventricular ejection fraction; LVIDd, left ventricular internal diameter–diastole; LVIDs, left ventricular internal diameter–systole; VUS, variants of uncertain significance.
*Denotes *P* value < 0.05.

Discussion

Clinically Actionable Findings in Diverse Biobank Participants

The utility of genetic information improves with deep and diverse genetic databases. This principle underlies All-of-Us and the Million Veteran programs, which aim to provide a broad genetic picture of the diverse US population. For hypertrophic cardiomyopathy, Manrai et al previously suggested that the underrepresentation of participants from diverse racial and ethnic backgrounds in large public databases led to the misclassification of benign variants as pathogenic, especially in populations of non-European ancestry.[25] In this study, individuals with pathogenic or likely pathogenic variants in the cardiac actionable genes had evidence of cardiac clinical findings in the electronic health record, and this was demonstrated by the 4 diverse biobank participants with variants in genes linked to long-QT syndrome. Because these variants increase risk for sudden death,[28] these genetic findings represent an opportunity for risk reduction. These findings differ from a previous report of rare long-QT syndrome–associated variants in individuals studied using electronic health records.[29] However, this current analysis relied more on ClinVar for interpretation, and during the intervening 3 years since the prior study, ClinVar has expanded with data contributions from additional clinical testing and now includes many more variants interpreted through more consistent guidelines.

Racial Differences in Unreported Variants

Given its composition, ClinVar's catalog of genetic variation, in part, reflects genetic testing practices.[5] The findings that people of European ancestry are less likely to have unreported variants than other race/ethnic groups ($P<0.001$) may reflect that clinical genetic testing is disproportionally applied to this group. However, this observation also likely reflects the smaller degree of genetic diversity within this group. Genetic diversity may contribute to disparity in interpreting genetic testing results in individuals of non-European, and especially African, ancestry.[30,31] This study showed that VUSs were disproportionally higher in individuals of African ancestry than all other individuals, and this trend continued when the cardiac medical actionable genes were analyzed, suggesting a potential source of disparity. These findings highlight the complexity of adjudicating variants in individuals of non-European ancestry because the high numbers of rare or private variants found in these populations are positioned to contribute to interpretation as a VUS.

VUSs and Echocardiographic Findings

This study identified echocardiographic findings associated with VUSs, suggesting that some VUSs may influence cardiac phenotype. LV internal diameters, in diastole and systole, correlated with having ≥1 VUS, and this result, which was seen in the entire cohort, appears to be driven by those carrying a cardiomyopathy diagnosis in the electronic health

record. A limitation of this study is that clinical correlates relied on electronic health record data, which are restricted by the subject's participation in the healthcare system and physician practice patterns. For example, the average duration between imaging studies was ≈5 years (SD, 3 years). Assessing the role of any of these variants more fully will require larger population-based studies, preferably conducted in a prospective manner over a much longer time interval. As a more focused analysis, family-based segregation studies, including clinical evaluation, can help resolve VUS status. Both approaches underscore the need for additional investigation in this area.

An array of cardiomyopathy genes had VUSs in these participants, and *MYBPC3* was noted as having the highest number of VUSs. *MYBPC3* truncations are seen in hypertrophic cardiomyopathy in which haploinsufficiency is thought to cause disease, and are therefore more likely to be interpreted as pathogenic or likely pathogenic than missense variants. Correspondingly, rare missense *MYBPC3* variation is more likely to be interpreted as VUS.[1,32] The *MYBPC3* VUSs in this current study were all missense and were scored as damaging or probably damaging to protein function by the in silico tool PolyPhen-2, suggesting that such variants may alter protein interactions or protein folding and stability; however, functional studies are needed to examine any such effect. SHaRe (Sarcomeric Human Cardiomyopathy Registry) reported that individuals with hypertrophic cardiomyopathy with pathogenic/likely pathogenic sarcomere mutations had a 2-fold greater risk of adverse outcomes than subjects with hypertrophic cardiomyopathy but without sarcomere mutations; those with a VUS in a sarcomere gene had intermediate risk.[32] These data support that VUSs may impart phenotype and highlight the need for refinements in variant interpretation. In the context of genetic testing for cardiomyopathies, VUSs are typically returned and can sometimes be interpreted with further familial testing and segregation analysis.[33,34] In the setting of biobank testing, such results would not be returned to subjects or providers, in part because interpretation of pathogenic variation is done outside the context of phenotype. We observed the correlation of VUSs with LV dimensions when viewing the entire cohort, but this finding was evident when considering only those with a cardiomyopathy diagnosis, suggesting a need to interpret variants in the presence of phenotype and family history. The identification of genetic variants contributing to cardiomyopathy affects management, especially for the accompanying arrhythmia risk.[1] The inability to fully interpret these variants limits the use of these data for both the patients and their family members. Improved methods in which variants are interpreted in concert with clinical diagnoses may address this deficiency.

An uptick in genetic testing in the diverse clinical setting, along with stricter guidelines on interpretation, and the limitations of in silico tools have contributed to an increased number of uncertain variants.[35,36] In practice, the enrichment of VUSs in specific racial groups makes genetic testing harder to interpret within those groups. Expanding genetic databases to include self-reported race and ultimately linking these data to health information should facilitate genetic interpretation. The data set developed in this report, along with the additional data generated from the eMERGE consortium, extends the diversity of publicly available genetic information.[16] Until databases are sufficiently powered to address these deficiencies, in some cases, it may be reasonable to return VUS results to biobank participants, especially if there are correlating clinical diagnoses. Return of any result should respect the biobank participant's wish to receive results. This would allow participants and their healthcare providers to assess risk by integrating genetic data with personal medical findings and family medical history.

Acknowledgments

We thank the McDonnell Genome Institute at Washington University in St. Louis. We also gratefully acknowledge the participation of the NUgene biobank participants.

Sources of Funding

This work was supported by grants from the National Institutes of Health (NIH U01HG008673, Chisholm; NIH R01HL128075, McNally; NIH/NLM T32 LM012203, Pottinger; NIH/NIDDK T32 DK007169, Pottinger), and the American Heart Association (18CDA34110460, Puckelwartz). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the article.

Disclosures

McNally is a consultant to Invitae Corp and Tenaya Therapeutics. The remaining authors have no disclosures to report.

References

1. Hershberger RE, Givertz MM, Ho CY, Judge DP, Kantor PF, McBride KL, Morales A, Taylor MRG, Vatta M, Ware SM. Genetic evaluation of cardiomyopathy: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2019;21:2406–2409.

2. Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, Horick N, Yang S, Shannon KM, Tung N, Ford JM, Lincoln SE, Ellisen LW. Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. *JAMA Oncol*. 2015;1:943–951.

3. Riley BD, Culver JO, Skrzynia C, Senter LA, Peters JA, Costalas JW, Callif-Daley F, Grumet SC, Hunt KS, Nagy RS, McKinnon WC, Petrucelli NM, Bennett RL, Trepanier AM. Essential elements of genetic cancer risk assessment, counseling, and testing: updated recommendations of the National Society of Genetic Counselors. *J Genet Couns*. 2012;21:151–161.

4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of

the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.

5. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44: D862–D868.
6. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15:565–574.
7. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2. 0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2016;19:249.
8. Bookman EB, Langehorne AA, Eckfeldt JH, Glass KC, Jarvik GP, Klag M, Koski G, Motulsky A, Wilfond B, Manolio TA, Fabsitz RR, Luepker RV; NHLBI Working Group. Reporting genetic results in research studies: summary and recommendations of an NHLBI working group. *Am J Med Genet Part A*. 2006;140:1033–1040.
9. Jarvik GP, Amendola LM, Berg JS, Brothers K, Clayton EW, Chung W, Evans BJ, Evans JP, Fullerton SM, Gallego CJ, Garrison NA, Gray SW, Holm IA, Kullo IJ, Lehmann LS, McCarty C, Prows CA, Rehm HL, Sharp RR, Salama J, Sanderson S, Van Driest SL, Williams MS, Wolf SM, Wolf WA; eMERGE Act-ROR Committee and CERC Committee; CSER Act-ROR Working Group, Burke W. Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *Am J Hum Genet*. 2014;94:818–826.
10. National Heart L, Blood Institute working g, Fabsitz RR, McGuire A, Sharp RR, Puggal M, Beskow LM, Biesecker LG, Bookman E, Burke W, Burchard EG, Church G, Clayton EW, Eckfeldt JH, Fernandez CV, Fisher R, Fullerton SM, Gabriel S, Gachupin F, James C, Jarvik GP, Kittles R, Leib JR, O'Donnell C, O'Rourke PP, Rodriguez LL, Schully SD, Shuldiner AR, Sze RKF, Thakuria JV, Wolf SM, Burke GL. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. *Circulation Cardiovasc Genet*. 2010;3:574–580.
11. Slavin TP, Van Tongeren LR, Behrendt CE, Solomon I, Rybak C, Nehoray B, Kuzmich L, Niell-Swiller M, Blazer KR, Tao S, Yang K, Culver JO, Sand S, Castillo D, Herzog J, Gray SW, Weitzel JN. Prospective study of cancer genetic variants: variation in rate of reclassification by ancestry. *J Natl Cancer Inst*. 2018;110:1059–1066.
12. Landry LG, Rehm HL. Association of racial/ethnic categories with the ability of genetic tests to detect a cause of cardiomyopathy. *JAMA Cardiol*. 2018;3:341–345.
13. Caswell-Jin JL, Gupta T, Hall E, Petrovchich IM, Mills MA, Kingham KE, Koff R, Chun NM, Levonian P, Lebensohn AP, Ford JM, Kurian AW. Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. *Genet Med*. 2018;20:234–239.
14. Ormond KE, Cirino AL, Helenowski IB, Chisholm RL, Wolf WA. Assessing the understanding of biobank participants. *Am J Med Genet Part A*. 2009;149A:188–198.
15. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, Struewing JP, Wolf WA; eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13.
16. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, Sanderson SC, Kannry J, Zinberg R, Basford MA, Brilliant M, Carey DJ, Chisholm RL, Chute CG, Connolly JJ, Crosslin D, Denny JC, Gallego CJ, Haines JL, Hakonarson H, Harley J, Jarvik GP, Kohane I, Kullo IJ, Larson EB, McCarty C, Ritchie MD, Roden DM, Smith ME, Bottinger EP, Williams MS; eMERGE Team.The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. 2013;15:761–771.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.
18. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
19. Puckelwartz MJ, Pesce LL, Nelakuditi V, Dellefave-Castillo L, Golbus JR, Day SM, Cappola TP, Dorn GW, Foster IT, McNally EM. Supercomputing for the parallelization of whole genome analysis. *Bioinformatics*. 2014;30:1508–1513.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38: e164.
21. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*. 1936;8:3–62.
22. Hsu J, Pacheco JA, Stevens WW, Smith ME, Avila PC. Accuracy of phenotyping chronic rhinosinusitis in the electronic health record. *Am J Rhinol Allergy*. 2014;28:140–144.
23. Bowen MS, Kolor K, Dotson WD, Ned RM, Khoury MJ. Public health action in genomics is now needed beyond newborn screening. *Public Health Genomics*. 2012;15:327–334.
24. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337:64–69.
25. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, Margulies DM, Loscalzo J, Kohane IS. Genetic misdiagnoses and the potential for health disparities. *New Eng J Med*. 2016;375:655–665.
26. Tarailo-Graovac M, Zhu JYA, Matthews A, van Karnebeek CDM, Wasserman WW. Assessment of the ExAC data set for the presence of individuals with pathogenic genotypes implicated in severe Mendelian pediatric disorders. *Genet Med*. 2017;19:1300–1308.
27. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM. A global reference for human genetic variation. *Nature*. 2015;526.
28. Bezzina Connie R, Lahrouchi N, Priori Silvia G. Genetics of sudden cardiac death. *Circ Res*. 2015;116:1919–1936.
29. Van Driest SL, Wells QS, Stallings S, Bush WS, Gordon A, Nickerson DA, Kim JH, Crosslin DR, Jarvik GP, Carrell DS, Ralston JD, Larson EB, Bielinski SJ, Olson JE, Ye Z, Kullo IJ, Abul-Husn NS, Scott SA, Bottinger E, Almoguera B, Connolly J, Chiavacci R, Hakonarson H, Rasmussen-Torvik LJ, Pan V, Persell SD, Smith M, Chisholm RL, Kitchner TE, He MM, Brilliant MH, Wallace JR, Doheny KF, Shoemaker MB, Li R, Manolio TA, Callis TE, Macaya D, Williams MS, Carey D, Kapplinger JD, Ackerman MJ, Ritchie MD, Denny JC, Roden DM. Association of arrhythmia-related genetic variants with phenotypes documented in electronic medical records. *JAMA*. 2016;315:47–57.
30. Gerhard GS, Fisher SG, Feldman AM. Genetic testing for inherited cardiac diseases in underserved populations of non-european ancestry: double disparity. *JAMA Cardiol*. 2018;3:273–274.
31. Landry LG, Ali N, Williams DR, Rehm HL, Bonham VL. Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Aff (Millwood)*. 2018;37:780–785.
32. Ho CY, Day SM, Ashley EA, Michels M, Pereira AC, Jacoby D, Cirino AL, Fox JC, Lakdawala NK, Ware JS, Caleshu CA, Helms AS, Colan SD, Girolami F, Cecchi F, Seidman CE, Sajeev G, Signorovitch J, Green EM, Olivotto I. Genotype and lifetime burden of disease in hypertrophic cardiomyopathy: insights from the sarcomeric human cardiomyopathy registry (SHaRe). *Circulation*. 2018;138:1387–1398.
33. Wolf MJ, Noeth D, Rammohan C, Shah SH. Complexities of genetic testing in familial dilated cardiomyopathy. *Circ Cardiovasc Genet*. 2016;9:95–99.
34. Dellefave L, McNally EM. The genetics of dilated cardiomyopathy. *Curr Op Cardiol*. 2010;25:198–204.
35. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol*. 2017;18:225.
36. Bean LJH, Hegde MR. Clinical implications and considerations for evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Med*. 2017;9:111.

SUPPLEMENTAL MATERIAL

Table S1. Medically Actionable Genes by Cardiac and Cancer Designation.

Cardiac Genes	Cancer Genes	Other Genes
ACTA2	*APC*	*ATP7B*
ACTC1	*BMPR1A*	*CACNA1S*
APOB	*BRCA1*	*OTC*
COL3A1	*BRCA2*	*RYR1*
DSC2	*MEN1*	
DSG2	*MLH1*	
DSP	*MSH2*	
FBN1	*MSH6*	
GLA	*MUTYH*	
KCNH2	*NF2*	
KCNQ1	*PMS2*	
LDLR	*PTEN*	
LMNA	*RB1*	
MYBPC3	*RET*	
MYH11	*SDHAF2*	
MYH7	*SDHB*	
MYL2	*SDHC*	
MYL3	*SDHD*	
PCSK9	*SMAD4*	
PKP2	*STK11*	
PRKAG2	*TP53*	
RYR2	*TSC1*	
SCN5A	*TSC2*	
SMAD3	*VHL*	
TGFBR1	*WT1*	
TGFBR2		
TMEM43		
TNNI3		
TNNT2		
TPM1		

Table S2. Pair-wise comparison of the number of VUS/person across groups (Figure 2 expanded)
All p values adjusted for multiple comparisons

ClinVar	African	European	Hispanic	Mixed
Nonsynonymous Coding Variation				
African	-	p<0.0001	p<0.0001	p=0.0023
European	-	-	ns	p=0.0027
Hispanic	-	-	-	p=0.0026
59 ACMG Actionable Genes				
African	-	p=0.00089	ns	ns
European	-	-	ns	ns
Hispanic	-	-	-	ns
30 ACMG Actionable Cardiac Genes				
African	-	p=0.0042	p=0.051	p=0.0061
European	-	-	ns	ns
Hispanic	-	-	-	ns

Table S3. Longitudinal Association of Left Ventricular Measures with the number of VUS/person in the Cardiac Actionable Genes in the entire cohort.

Echo Measure VUS ≥1 vs VUS=0	**Total Observations**	**African**	**European**	**Hispanic**	**Mixed**
LVIDd/BSA	n=914	n=266	n=316	n=151	n=181
Slope Difference Estimate	**+0.018 (p=0.042)**	+0.0021	**+0.034 (p=0.038)**	+0.042 (p=0.093)	+0.025
LVIDs/BSA	n=984	n=277	n=327	n=169	n=211
Slope Difference Estimate	**+0.027 (p=0.0016)**	**+0.026 (p=0.029)**	**+0.047 (p=0.013)**	+0.027	+0.0058
LVEF	n=987	n=273	n=319	n=166	n=229
Slope Difference Estimate	-0.33	**-0.85 * (p=0.046)**	-0.32	+0.036	+0.36
IVSDd/BSA	n=1,049	n=296	n=341	n=185	n=227
Slope Difference Estimate	-0.00068	-0.0063	+0.0092 (p=0.056)	-0.0063	-0.0055

† p<0.1
Slope difference estimates compared 1 or more VUS to 0 VUS, and the model controlled for age at echocardiogram, sex, and self- reported race/ethnicity.
LVIDd=left ventricular diameter diastole; BSA=body surface area; IVSDd=interventricular septal end diastole
LVIDs=left ventricular diameter systole; LVEF=left ventricular ejection fraction

Table S4. Longitudinal Association of Left Ventricular Measures with VUS in Cardiac Actionable Genes from 94 Subjects with Cardiomyopathy Diagnostic Codes by Genetic Race/Ethnicity

Echo Measure VUS ≥1 vs VUS=0	**Total Observations**	**African**	**European**	**Hispanic**	**Mixed**
LVIDd/BSA	n=447	n=138	n=238	n=45	n=26
Slope Difference Estimate	**+0.035 (p=0.0035)**	+0.019	**+0.038 (0.044)**	+0.076	-
LVIDs/BSA	n=442	n=130	n=239	n=46	n=27
Slope Difference Estimate	**+0.047 (p=0.0003)**	**+0.033 (p=0.029)**	**+0.05 (0.023)**	+0.068	**+0.1 (p=0.011)**
LVEF	n=426	n=131	n=225	n=44	n=26
Slope Difference Estimate	-0.42	-0.086	-0.25	-0.44	+9.01 (p=0.052)
IVSDd/BSA	n=465	n=141	n=248	n=48	n=28
Slope Difference Estimate	-0.0014	-0.0094	+0.008	-	-

Slope difference estimates compared ≥1 VUS to 0 VUS, and the model controlled for age at echocardiogram, sex, and genetic race/ethnicity
LVIDd=left ventricular diameter diastole; BSA=body surface area; IVSDd=interventricular septal end diastole
LVIDs=left ventricular diameter systole; LVEF=left ventricular ejection fraction

Table S5. Cardiomyopathy Genes.

A2ML1	*CRYAB*	*FKRP*	*LAMP2*	*MYOZ*	*RAF1*	*TGFB3*
ABCC9	*CSRP3*	*FKTN*	*LDB3*	*MYPN*	*RASA1*	*TMEM43*
ACTC1	*CTF1*	*FLNC*	*LMNA*	*NEBL*	*RBM20*	*TMPO*
ACTN2	*CTNNA3*	*GAA*	*LRRC10*	*NEXN*	*RIT1*	*TNNC1*
AGL	*DES*	*GATA4*	*MAP2K1*	*NF1*	*RRAS*	*TNNI3*
ALPK3	*DMD*	*GATA6*	*MAP2K2*	*NKX2-5*	*RYR2*	*TNNT2*
ANKRD1	*DOLK*	*GATAD1*	*MIB1*	*NPPA*	*SCN5A*	*TPM1*
BAG3	*DSC2*	*GLA*	*MURC*	*NRAS*	*SGCD*	*TRDN*
BRAF	*DSG2*	*HCN4*	*MYBPC3*	*PDLIM3*	*SHOC2*	*TTN*
CACNA1C	*DSP*	*HRAS*	*MYH6*	*PKP2*	*SLC22A5*	*TTR*
CALR3	*DTNA*	*ILK*	*MYH7*	*PLEKHM2*	*SOS1*	*TXNRD2*
CASQ2	*EMD*	*JPH2*	*MYL2*	*PLN*	*SOS2*	*VCL*
CAV3	*EYA4*	*JUP*	*MYL3*	*PRDM16*	*SPRED1*	
CBL	*FHL1*	*KRAS*	*MYLK2*	*PRKAG2*	*TAZ*	
CHRM2	*FHL2*	*LAMA4*	*MYOM1*	*PTPN11*	*TCAP*	

Gene list ascertained by combining commercially available gene testing panels.

Table S6. Variants of uncertain significance found in participants with a cardiomyopathy diagnosis.

Cardiac Actionable Genes w/VUS	Variant	gnomAD allele frequency	Cardio-myopathy Genes w/ Pathogenic Variants	Variant	gnomAD Allele frequency
APOB	p.I2850M				
APOB	p.R297C				
APOB	p.G724C				
COL3A1	p.V976A				
DSC2	p.G8V	9.15e-5			
DSC2	p.T268A	1.38e-4			
DSC2	p.G755V	6.02e-5			
FBN1	p.H1130R	1.41e-5			
MYBPC3 ***LMNA***	p.L199P p.R401H	0 8.55e-5			
MYBPC3 ***MYL3***	p.G1248R p.R94C	3.21e-5 2.39e-5			
MYBPC3	p.C1266Y	0			
MYBPC3	p.R17Q	8.48e-5			
MYBPC3	p.R574Q	6.48e-5			
PCSK9	p.E57K		**TTR**	p.V122I	1.54e-3
PCSK9	c.42_43insCTGCTGCTG				
PRKAG2	p.R84W	4.78e-4			
PRKAG2	p.E185V	4.37e-5			
PRKAG2	p.T174M	7.42e-5			
RYR2	p.P466A	1e-4			
RYR2	p.R1055H	0			
SCN5A	p.S866L	2.48e-5			
SCN5A	p.A1100V	5.4e-5			
TGFBR1	c.50_51insGGCGGC	LCR			
TGFBR1	c.50_51insGGCGGC	LCR			

Genes associated with **cardiomyopathies** are shown **bolded**. Genes not known to be associated with a myopathic condition are shown in light grey, as these are less likely to contribute directly to changes in LV dimensions. Genes associated with some myopathic conditions are shown in normal text coloring. LCR indicates the variant derives from a low complexity region, frequency difficult to ascertain. TTRp.V122I is also known as p.V142I.

Table S7. Additional Information on VUSs found in cardiomyopathy subjects.

Gene	Variant	Frequency	Diagnosis Listed in ClinVar	Most recent ClinVar report
DSC2	p.G8V	9.15e-5	Cardiomyopathy	2018
DSC2	p.T268A	1.38e-4	Cardiomyopathy, ARVC	2018
DSC2	p.G755V	6.02e-5	ARVC	2017
FBN1	p.H1130R	1.41e-5	Marfan,TAAD	2018
MYBPC3 ***LMNA***	p.L199P p.R401H	0 8.55e-5	HCM CMT	2015 2018
MYBPC3 ***MYL3***	p.G1248R p.R94C	3.21e-5 2.39e-5	Cardiomyopathy, HCM Cardiomyopathy	2018 2016
MYBPC3	p.C1266Y	0	Cardiomyopathy, HCM	2018
MYBPC3	p.R17Q	8.48e-5	Cardiomyopathy, HCM	2018
MYBPC3	p.R574Q	6.48e-5	HCM, DCM, LVNC	2018
PRKAG2	p.R84W	4.78e-4	Cardiomyopathy	2018
PRKAG2	p.E185V	4.37e-5	Glycogen storage heart	2018
PRKAG2	p.T174M	7.42e-5	Glycogen storage heart	2018
RYR2	p.P466A	1e-4	Ventricular tachycardia	2018
RYR2	p.R1055H	0	Ventricular tachycardia	2018
SCN5A	p.S866L	2.48e-5	Brugada	2018
SCN5A	p.A1100V	5.4e-5	Cardiovascular phenotype, LQTS, Brugada	2018
TTR	p.V122I (AKA, p.V142I) (PATH)	1.54e-3	Carpal Tunnel, Transthyretin Cardiomyopathy	2018

ARVC, Arrhythmogenic Right Ventricular Cardiomyopathy
CMT, Charcot Marie Tooth
TAAD, Thoracic Aortic Aneurysm Dissection
LVNC, Left Ventricular Noncompaction Cardiomyopathy
MYBPC3, p.G1247R = p.G1248R

Table S8. Pair-wise comparison of the number of Unreported Variants per person across groups.
All p values adjusted for multiple comparisons

ClinVar	African	European	Hispanic	Mixed
Nonsynonymous Coding Variation				
African	-	p<0.0001	p<0.0001	p<0.0001
European	-	-	p<0.0001	p<0.0001
Hispanic	-	-	-	p<0.0001
59 ACMG Actionable Genes				
African	-	ns	ns	ns
European	-	-	ns	ns
Hispanic	-	-	-	ns
30 ACMG Actionable Cardiac Genes				
African	-	p=0.0011	p=0.0014	ns
European	-	-	ns	p=0.048
Hispanic	-	-	-	p=0.066

Table S9. Longitudinal Association of Left Ventricular Measures with the number of Unreported Variants/person in Cardiac Actionable Genes in the entire cohort.

Echo Measure VUS ≥1 vs VUS=0	**Total Observations**	**African**	**European**	**Hispanic**	**Mixed**
LVIDd/BSA	n=914	n=266	n=316	n=151	n=181
Slope Difference Estimate	-0.069	-0.016	+0.0035	-0.028	+0.012
LVIDs/BSA	n=984	n=277	n=327	n=169	n=211
Slope Difference Estimate	+0.0058	+0.012	+0.019	-0.028	-0.0036
LVEF	n=987	n=273	n=319	n=166	n=229
Slope Difference Estimate	+0.35	+0.8 †	-0.3	+0.29	+0.47
IVSDd/BSA	n=1,049	n=296	n=341	n=185	n=227
Slope Difference Estimate	-0.0013	+0.0043	-0.0074	+0.0074	-0.0072

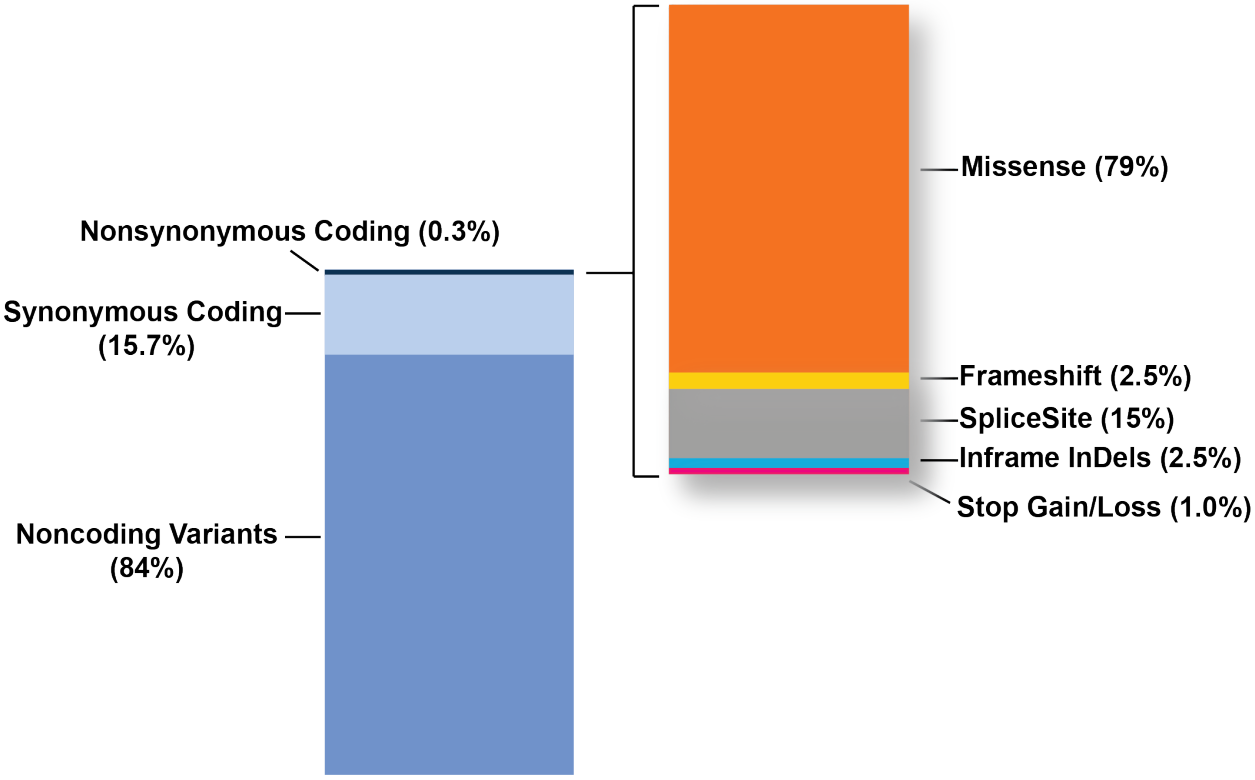
No values had $p<0.05$

Slope difference estimate compared ≥3 or more VUS to 2 VUS, and model controlled for age at echocardiogram, sex, and self-reported race/ethnicity

LVIDd=left ventricular diameter diastole; BSA=body surface area; IVSDd=interventricular septal end diastole

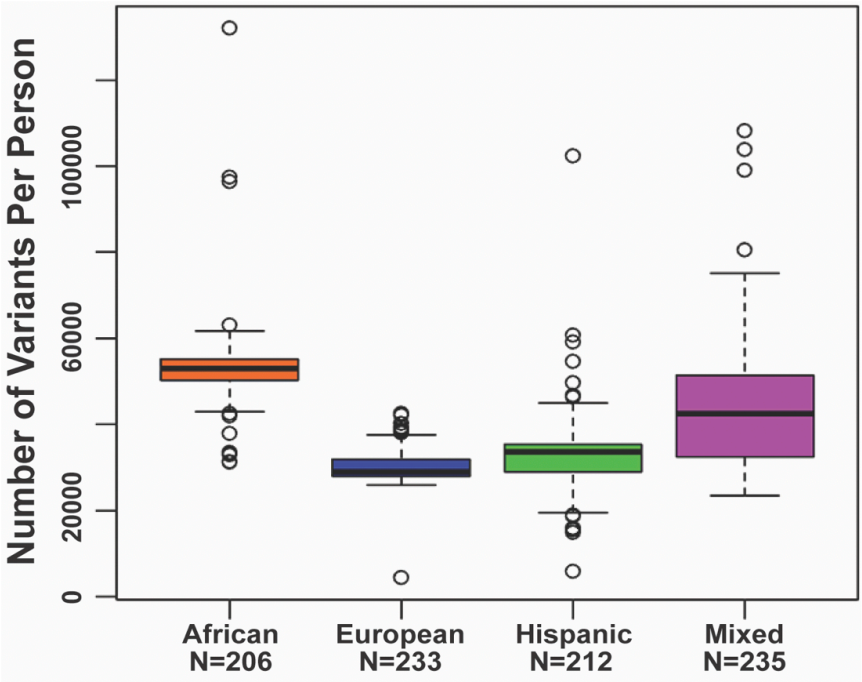
LVIDs=left ventricular diameter systole; LVEF=left ventricular ejection fraction

Figure S1. Distribution of genetic variation in the NUGene cohort as determined by whole genome sequencing.



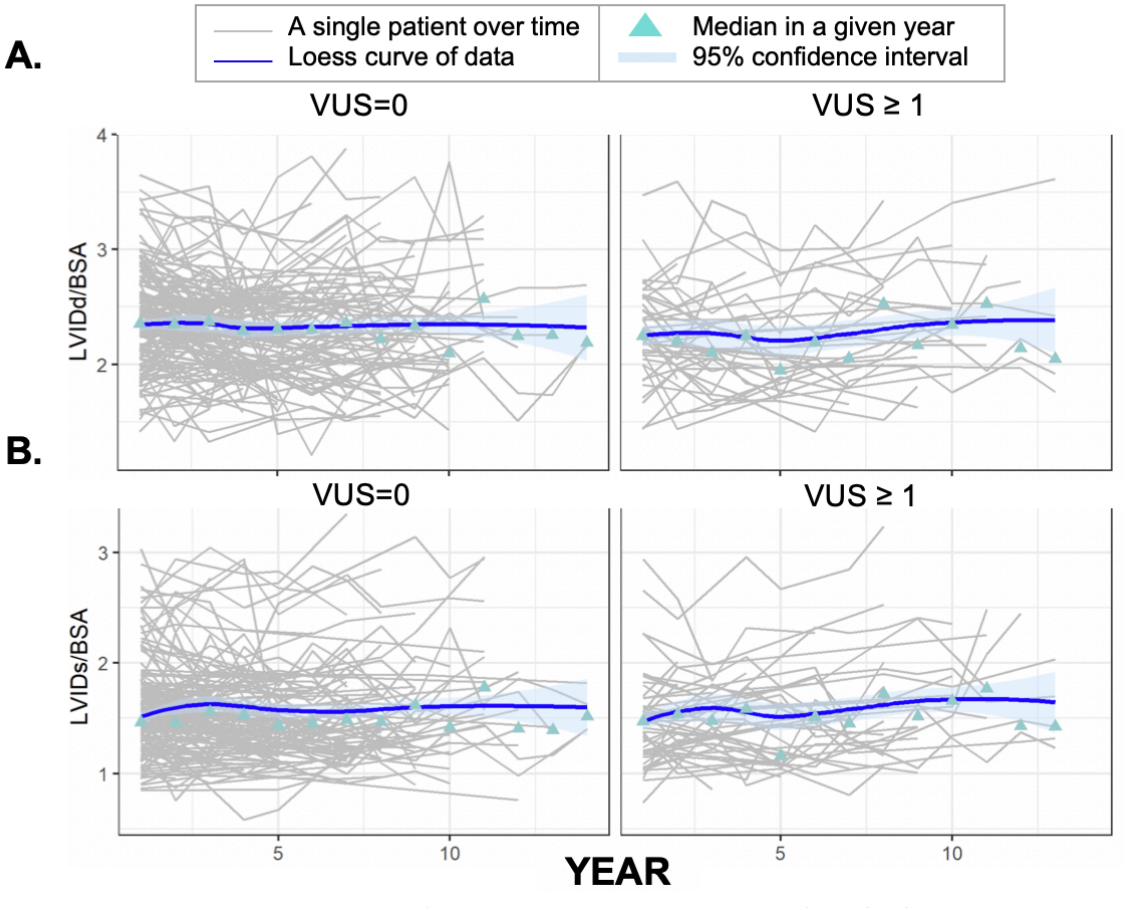
On average, 5,300,085±416,604 variants per person were identified in comparison the reference human genome. When restricting this analysis to nonsynonymous variants in the coding region, 17,282±1,234 variants per person were identified. Of these, per person there were 13,618±975 missense variants (orange), 400±31 frame shift variants (yellow), 2,500 ±195 splice site variants (grey); 400±35 inframe insertions or deletions (in/dels, light blue), and 200±13 variants that introduced a new stop codon or removed a stop codon (stop gain/ loss, pink).

Figure S2. The number of variants observed only once within NUgene populations.



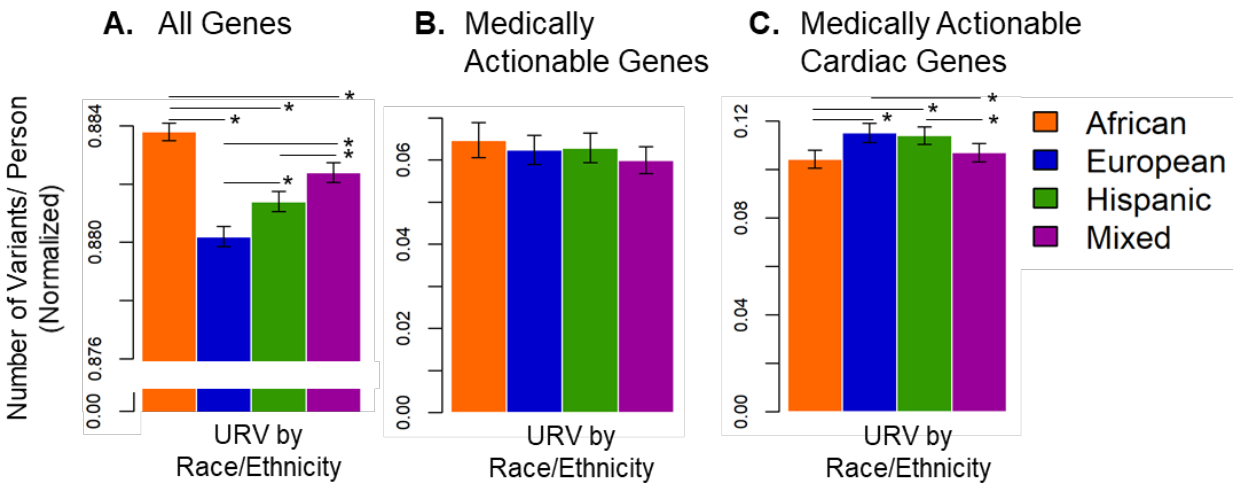
The average number of variants that are unique to individuals in the cohort by self-reported race/ethnicity; African 52,865±8,320; European 30,217±3,992; Hispanic 32,664±7,851; and mixed race 43,513±13,199. All values are per person ($p < 0.0001$ ANOVA across all self-reported race/ethnicities.)

Figure S3. Median left ventricular dimensions correlated to VUS number in cardiac actionable genes across the entire cohort.



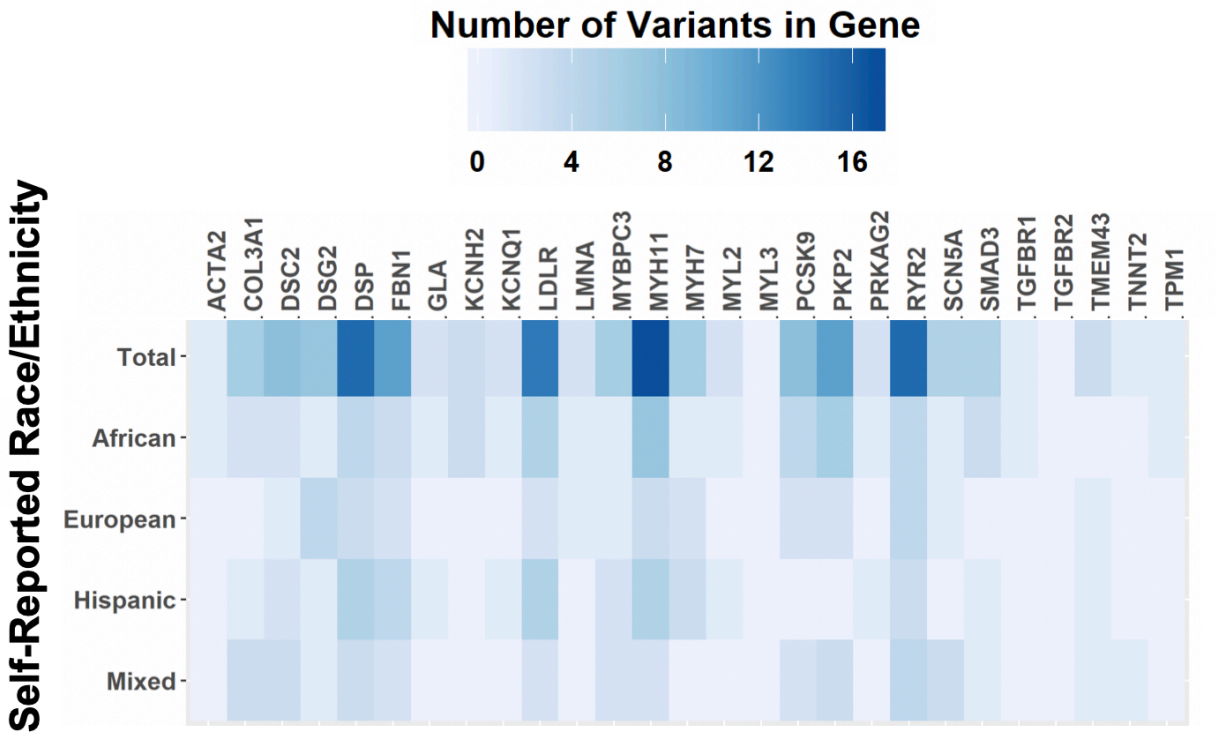
(A) Left ventricular internal dimension in diastole (LVIDd) corrected for body surface area (BSA) over 14 years of hospital visits by the number of variants of uncertain significance found in the cardiac actionable genes. **(B)** Left ventricular internal dimension in systole (LVIDs) corrected for BSA over 14 years of hospital visits by the number of variants of uncertain significance found in the cardiac actionable genes.

Figure S4. African ancestry biobank participants have significantly more unreported variants (URV).



(A) The average number of total coding variants across all genes is shown for each group based on self-reported race/ethnicity. Unreported variants (URV) in ClinVar were greater in African ancestry biobank participants compared to other groups ($p<0.0001$, ANOVA across all groups). **(B)** The average number of coding variants in the 59 medically actionable genes is shown by self-reported race/ethnicity. The number of unreported variants (URV) in ClinVar were not different across groups. **(C)** The average number of coding variants in the 30 cardiac actionable genes is shown by self-reported race/ethnicity. The number of unreported variants (URV) in ClinVar were less in African ancestry biobank participants compared to all other groups excluding Hispanics ($p<0.0001$). Pairwise comparison exact p-values are in Table S9.

Figure S5. Unreported Variants in medically actionable cardiac genes by Race/Ethnicity.



The number of unreported variants in cardiac actionable genes in the NUgene cohort. *APOB* is not shown.